

REVISTA ESPAÑOLA DE DOCUMENTACIÓN CIENTÍFICA
30, 2, ABRIL-JUNIO, 165-179, 2007
ISSN 0210-0614

ESTUDIOS / RESEARCH STUDIES

La identificación temática a partir de la visualización de la información: una aproximación mediante el caso de *women* en Medline*

Topic identification based on information visualization: an approximation through the case of women in Medline

Benjamín Vargas-Quesada, Irene Doménech Coullaut, Gema García Gómez, Carmen Sánchez Domínguez, Ana Extremeño Placer, María Ángeles Zulueta García**

Resumen: La salud de las mujeres es un aspecto de gran interés que surge con fuerza en la década de los 90. El objetivo de este trabajo consiste en delimitar un tema de gran actualidad y relevancia como es la investigación sobre la mujer, desde el punto de vista de las ciencias de la salud, reflejada en las publicaciones científicas. Consideramos que la mejor fuente para este objetivo es Medline. Para ello, desarrollamos una metodología en tres niveles consecutivos que nos permite poner de manifiesto la estructura intelectual del concepto de mujer en Medline, así como detectar e identificar sus principales descriptores en la base de datos seleccionada.

Palabras clave: Medline, mujer, visualización de información, identificación temática.

Abstract: Women's health is an aspect of great interest that arises with force in the 90s. The objective of this work consists in delimiting a topic of great interest and relevance as is the investigation on women from the point of view of health sciences, reflected in scientific publications. We considered that the best source for this objective is Medline. To this end we developed a methodology in three consecutive levels that allows us to show the intellectual structure of the concept of women in Medline, as well as to detect and identify its main descriptors in the database selected.

Keywords: Medline, women, information visualization, topic identification.

* Una versión preliminar de este trabajo fue presentada en el III Seminario Internacional sobre estudios cuantitativos y cualitativos de la ciencia y la tecnología. Congreso Internacional de Información INFO 2006. La Habana, Cuba, 17-21 abril, 2006.

** Facultad de Documentación. Universidad de Alcalá. Alcalá de Henares (Madrid).
Correo-e: ma.zulueta@uah.es.
Recibido: 23-10-2006; 2.^a versión: 8-3-2007.

1 Introducción

Los análisis bibliométricos ofrecen una imagen muy detallada del estado de la investigación en las diferentes áreas científicas, disciplinas o materias. Uno de los principales problemas que surgen en el momento de delimitar los contenidos específicos de una investigación procede del sistema o de los criterios utilizados para ello. Hoy en día, existe un gran número de sistemas de clasificación como la CDU, encabezamientos de materias, clasificación de revistas y descriptores que facilitan esta tarea. La utilización de descriptores aporta una gran ventaja, ya que permite identificar perfectamente todos o la mayor parte de los contenidos intelectuales de un documento científico, con una precisión mayor que los otros métodos. Una de las principales bases de datos que utiliza este sistema es Medline (National Library of Medicine, 2005a). En esta base de datos hay materias que están perfectamente identificadas por medio de descriptores específicos. La mayoría de los estudios que han utilizado los descriptores del *Medical Subject Heading* (MeSH) (National Library of Medicine, 2006) para la delimitación temática de una materia (Lin, 2005), se han limitado al empleo de un único descriptor. No obstante, a veces es necesario delimitar un tema para el que no existe un descriptor concreto y hay que recurrir a un conjunto de descriptores. En este trabajo pretendemos delimitar un tema de gran actualidad y relevancia, como es la investigación sobre la mujer desde el punto de vista de las ciencias de la salud, reflejada en las publicaciones científicas. Este tema abarca enfoques muy variados como son los de tipo social, cultural, económico y, desde luego sanitario, que no pueden ser abarcados con un solo descriptor.

La salud de las mujeres es un aspecto de gran interés que surge con fuerza durante la década de los 90 del siglo pasado. Las políticas sanitarias reconocen las diferencias sociales y clínicas entre los hombres y las mujeres, y muchos países han aprobado acuerdos internacionales que recogen el papel determinante que el género tiene sobre la salud de los ciudadanos pero que, sin embargo, no se han plasmado en políticas concretas de ámbito nacional (Dirección General de Salud Pública, 2005). Por otra parte, también se ha empezado a plantear la introducción del concepto de género en la salud. Hasta ahora, la presencia de las mujeres como sujetos de investigación en aquellas enfermedades que afectaban indistintamente a hombres y mujeres era escasa. Pero, a partir de los años 90, comienzan a surgir iniciativas por parte de algunos gobiernos e instituciones para que la salud de los hombres y de las mujeres reciba la misma atención y se empiece a asimilar que las diferencias biológicas entre los sexos pueden influir en la severidad e incidencia de las enfermedades (Wizemann, 2001). Entre las iniciativas realizadas al respecto, se encuentran las del *National Institutes of Health* (NIH) que elaboró medidas para incluir a las mujeres como sujetos de análisis en las pruebas clínicas (NIH, Revitalization Act, 1993) y creó la *Office of Research on Women's Health* en 1990, para controlar y asegurar la investigación sobre salud y mujeres. Otros organismos e instituciones relacionados con la salud, también empiezan a introducir cambios y a elaborar instrumentos para motivar a los países a introducir la perspectiva del género y el sexo en sus políticas

sanitarias (Women's Health Bureau, 2002; Gender and Health Group at the Liverpool School of Tropical Medicine, 2000; WHO, 2002; Sen, 2005). Como consecuencia, las publicaciones sobre investigaciones que empiezan a considerar la perspectiva de la mujer han crecido considerablemente en muy poco tiempo y han generado una gran cantidad de información y de nuevos conceptos.

En este trabajo pretendemos delimitar la investigación sobre la mujer desde el punto de vista de las ciencias de la salud. Para ello se ha planteado un estudio en tres niveles consecutivos y nuestra pretensión es que en el tercer nivel sea posible la recuperación de los documentos relacionados con la salud de las mujeres que estén incluidos en Medline. El hecho de que no haya descriptores específicos para el tema elegido y de que la combinación de descriptores y etiquetas no responda a los criterios de especificidad y precisión que nos hemos planteado, ha condicionado la decisión de establecer estos tres niveles.

En el primer nivel se establece un punto de partida para identificar los descriptores que mejor representan el concepto de mujer y salud en Medline. Para ello hemos recurrido al DeCS (Descriptores en Ciencias de la Salud) (BIREME, 2005), que nos indica cuales son los descriptores más apropiados para la búsqueda y recuperación de información, a través del MeSH, que hacen referencia al concepto «mujeres» (*women*). Estos descriptores serán los que nos permitirán adentrarnos en el segundo nivel del proyecto.

El segundo nivel consiste en la ampliación del número de descriptores significativos relacionados con la salud y los aspectos clínicos de las mujeres. Consideraremos aspectos clínicos propios de las mujeres a aquellos exclusivos o cuya población más afectada es femenina, tales como el cáncer de mama, la menopausia, el embarazo, la terapia hormonal sustitutiva, etc. La generación de mapas conceptuales a partir de la coocurrencia de los descriptores, en conjunción con la teoría de grafos y las medidas típicas del análisis de redes (Wasserman, 1998; Batagelj, 2003), se convierten en herramientas esenciales en este caso.

Por último, el tercer nivel de análisis nos permitirá identificar los descriptores más específicamente relacionados con la salud y los aspectos clínicos de la mujer, así como su corolario. En él, se ponen de manifiesto conceptos muy amplios que abarcan a toda la población, pero que muestran una especial relación con las mujeres o su padecimiento por parte de ellas, obligando a la inclusión de estas enfermedades como parte importante de los problemas sanitarios padecidos por las mujeres. Además, este tercer nivel puede ser utilizado como un elemento de retroalimentación metodológica, ya que identifica descriptores candidatos a ser utilizados para enriquecer y ampliar el punto de partida del que hablábamos en el primer nivel.

El objetivo de este trabajo consiste en desarrollar una metodología que nos permita visualizar, detectar e identificar los principales descriptores relacionados con el tema planteado, así como estudiar la procedencia o no de utilizar dicha metodología para poner de manifiesto la estructura intelectual de la base de datos seleccionada, y delimitar materias que una estructura rígida y jerárquica como la que tiene el MeSH, no puede reflejar.

2 Material y métodos

Para la visualización y puesta de manifiesto de la estructura científica/intelectual de los trabajos relacionados con el concepto *women*, seguimos el esquema desarrollado por Börner (2003).

2.1 Origen de los datos y procesamiento

Puesto que nuestro objetivo es detectar e identificar los aspectos clínicos y médicos de la mujer, consideramos que Medline es la base de datos más adecuada para nuestros fines. Medline contiene referencias bibliográficas y resúmenes de autor de más de 4.800 revistas biomédicas publicadas en los Estados Unidos y otros 70 países. Esta base de datos almacena alrededor de 12 millones de referencias desde mediados de los años 60. Su cobertura es internacional, pero recoge preferentemente revistas en inglés o con resumen en este idioma. Esta base de datos dispone de unas herramientas que facilitan su utilización y que la convierten en la base de datos más apropiada para realizar el tipo de análisis que pretendemos llevar a cabo. Entre ellas está el sistema de indización de los documentos, basado en la utilización de un vocabulario controlado de alto prestigio como es el MeSH, que confiere a la base de datos un alto nivel de precisión y exhaustividad. El tesoro MeSH es una de las herramientas documentales más exitosas. Está compuesto por 33.000 términos que se estructuran jerárquicamente en 15 categorías que cubren todos los aspectos sanitarios. Su utilización para este análisis permite poner de manifiesto los aspectos de la mujer relacionados con la salud, así como otros de tipo social, cultural, político, económico y psicológico fundamentalmente.

Para la descarga de documentos se ha decidido utilizar los descriptores relacionados con el concepto *women*, que el DeCS considera como los más adecuados para la búsqueda y recuperación de temas relacionados con las mujeres. Estos descriptores son: *Women*, *Women's health*, *Women's Health Services*, *Women's rights*, *Battered women*, *Women Working*, *Dentists women*, *Physicians women* y *Pregnant women*. Todos ellos se han combinado con las *check tags Human* y *Female*, con el fin de excluir los documentos relacionados con animales y que no lo estuviesen exclusivamente con las mujeres. Con fines meramente investigadores, el día 11 de enero de 2006 descargamos de la base de datos Medline por medio de *WebSpirs* 5.03 (Ovid Technologies, 2005), todos los documentos que cumplían la estrategia de búsqueda anteriormente expuesta. Dichos documentos fueron descargados a una base de datos bibliográfica y posteriormente exportados a una base de datos relacional. Este nuevo repositorio contiene 19.524 documentos (journal articles, review, historical article, editorial, letter, comment, news, case reports, clinical trials...), que han sido clasificados mediante 76.320 descriptores (3.065 distintos), y que se han publicado en 2.648 revistas diferentes entre los años 1964 y 2005.

2.2 Unidades de análisis

Las unidades de medida más comúnmente utilizadas para la representación de dominios científicos son: revistas, documentos, autores, términos de indización y palabras. Aunque recientemente, a este grupo se les han añadido también otras unidades más amplias como son las categorías ISI (Moya-Anegón, 2004).

Un aspecto a tener en cuenta relacionado con las unidades de análisis, es la cantidad de información de que se dispone y subsidiariamente, el tamaño total del dominio a representar. Si el número de variables o ítems con los que vamos a trabajar es muy reducido, será posible construir visualizaciones de unidades muy pequeñas, como palabras o descriptores. Si no es así, habrá que plantearse el uso de otras unidades de análisis mayores, como los documentos o los autores. Pero si la cantidad de información a tener en cuenta es elevada o muy elevada, habrá que recurrir a unidades de análisis que cuenten con la capacidad de concentrar unidades más pequeñas, como es el caso de las revistas que agrupan a documentos, autores y términos. Esta consideración no es nueva en el campo de la visualización de la información, y está originada por la limitación física que supone la representación de grandes cantidades de información en un espacio reducido.

Al igual que otros autores (Boyack, 2005; Samoylenko, 2006), consideramos que las revistas son unidades informativas lo suficientemente explícitas, como para ser utilizadas en representación de las distintas unidades informativas que aglutinan. No obstante, y para realizar análisis de granularidad más fina, recurrimos a los descriptores. La visualización de las relaciones existentes entre descriptores (Lin, 2005; Chen, 2005) extraídos de bases de datos como PubMed (National Library of Medicine, 2005b), suponen un punto de partida muy adecuado desde el que realizar el análisis de contenido de las principales y/o diferentes líneas de trabajo dentro de una materia. Con el fin de conseguir la mayor precisión posible, hemos recurrido únicamente a los descriptores principales (*major*), eliminando además sus calificadores y los descriptores geográficos.

Tanto las revistas como los descriptores, en combinación con las adecuadas técnicas de reducción del espacio y de representación de la información, nos permiten realizar representaciones estáticas, en distintos niveles, de las principales áreas de investigación relacionadas con la mujer.

2.3 Unidades de medida

Para cuantificar las relaciones existentes entre cada una de las revistas seleccionadas, utilizaremos el grado de coocurrencia de los descriptores de cada uno de los documentos que se han publicado en ellas, como medida de similaridad. No utilizaremos ningún tipo de medida de normalización en los valores. Es decir, trabajaremos con valores de coocurrencia puros o *raw data*, tal y como recomienda Leydesdorff (2006).

El resultado que se obtiene como consecuencia de aplicar esta unidad de medida es una matriz simétrica de 2.648 por 2.648 revistas, donde los valores de su diagonal, deliberadamente, los hemos transformado en ceros.

2.4 Diseño

Existe hoy día en el mercado una amplia variedad de paquetes informáticos destinados a la representación y visualización de datos bibliométricos. Nosotros hemos optado por *Pajek* (Batagelj, 2003; Batagelj, 2006).

Pajek es un software europeo gratuito para la visualización y análisis de grandes redes sociales, cuya forma de presentar los gráficos facilita su interpretación. Este *software* ha sido ampliamente utilizado por otros autores para generar visualizaciones estáticas (por ejemplo, White, 2003; Börner, 2003), con fines similares al nuestro.

2.4.1 Reducción del espacio

Son muy distintas y variadas las técnicas que se utilizan para la reducción del espacio n -dimensional. Solas o en combinación con otras, las más comunes han sido: Escalamiento Multidimensional (MDS), Clustering, Análisis Factorial (AF), *Self-Organizing Maps* (SOMs) y PathfinderNetwork (PFNET).

Representar en un plano el grafo resultante de la matriz de coocurrencia de las 2.648 revistas que componen nuestra base de datos es una ardua tarea. Pero aún lo es más, hacerlo de una forma inteligible de modo que el ojo humano pueda captar y transmitir al cerebro la información gráfica para que sea procesada. Por tanto, es necesario reducir el espacio n -dimensional y para ello recurrimos al Análisis Factorial (AF). El mismo problema ocurre con los descriptores. Para solventarlo, en este caso, recurriremos a representar sólo los 25 más prominentes. Del mismo modo y puesto que las relaciones entre las distintas unidades de análisis son muy abundantes, recurrimos a PFNET, que mediante los parámetros de poda $r = \infty$ y $q = n - 1$ permite eliminar las relaciones superfluas y preservar sólo las esenciales (Chen, 1998).

2.4.2 Distribución espacial

Pajek incorpora dos algoritmos destinados a generar visualizaciones estéticas, desde el punto de vista del ojo humano: Kamada (1989) y Fruchterman (1991). La simetría, la distribución uniforme de nodos, la longitud uniforme de los enlaces, la reducción del número de cruces de los mismos, etc., juegan un papel fundamental en la decantación por un algoritmo u otro. Aunque estudios como el de Brandenburg (1996), no detectan la preponderancia de un algoritmo sobre otro, la mayor parte de la comunidad científica opta por el algoritmo de *Kamada-Kawai*. Nosotros hemos decidido también utilizarlo.

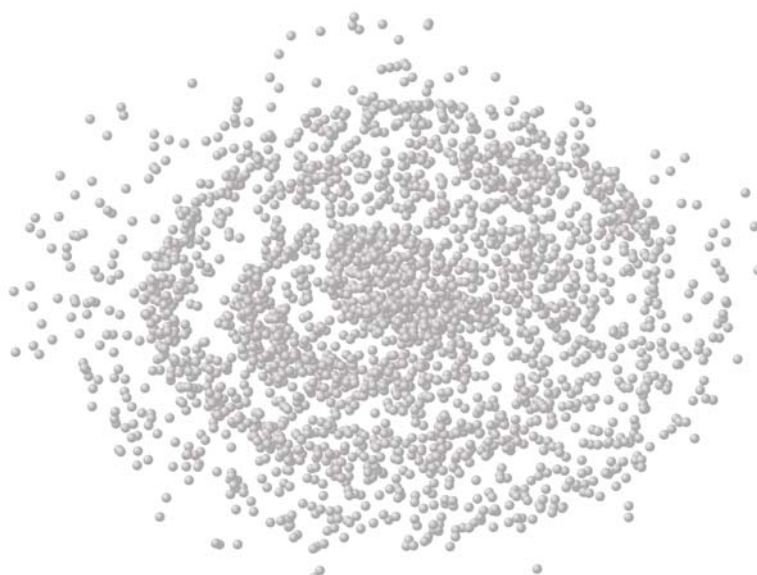
2.5 Visualización

De las distintas posibilidades que *Pajek* ofrece para exportar las representaciones con él diseñadas, hemos elegido *Scalable Vector Graphic* (SVG) (W3C, 2004). SVG es un formato ligero, veloz, ingenioso y gratuito que por ser vectorial, permite realizar *zoom in* y *zoom out*. Además, posibilita desplazamientos en cualquier dirección y la búsqueda de información en su interior. Cuenta también con un firme soporte técnico y un importante apoyo por parte de la industria del sector y de muchos expertos.

3 Resultados

La figura 1 es el grafo que se obtiene de la matriz de coocurrencia de las 2.648 revistas obtenidas. Para facilitar su visualización, hemos eliminado los enlaces o relaciones entre revistas, pues su número es tan elevado que imposibilita ver las posiciones que ocupan éstas. Esta imagen es la típica que se obtiene al representar un grafo mediante el algoritmo de *Kamada-Kawai*. Los nodos, en este caso revistas, que más se relacionan con el resto, tienden a situarse en la zona central, mientras que aquellos con menos relaciones o menos descriptores en común, se sitúan en la periferia.

Figura 1
Grafo de coocurrencia de revistas



Obviamente, identificar una revista en este conglomerado resulta prácticamente imposible. Aún más cuando se han omitido sus títulos, incluso los abreviados, ya que al mostrarlos se cubre la totalidad del gráfico, impidiendo la visualización de las esferas o nodos que representan a las revistas. Se hace necesario pues, algún método de reducción del espacio con el que aglutinar, identificar y representar la información, y este es el Análisis Factorial (AF).

El AF identifica nueve factores en la matriz de coocurrencias de 2.648 por 2.648 revistas, los cuales explican un 99% de la varianza total (tabla I). La figura 2 es el resultado de trasladar el AF al grafo anterior. Cada factor o área de investigación se ha coloreado de forma distinta, y se le ha asignado un nombre según el método propuesto por Moya-Anegón (1998): se han ordenado los factores por su índice de ponderación en orden decreciente y a partir de la coincidencia de palabras o conceptos de los títulos de revistas con un valor mayor o igual a 0,7, se le ha asignado un nombre a cada factor. Se han identificado nueve factores cuyo nombre y descripción son los siguientes: 1) Salud: Condición física y mental de las mujeres; 2) Profesión médica: Ejercicio de la profesión médica por mujeres y aspectos relacionados con su formación; 3) Medicina general e interna: Diagnóstico y tratamiento de enfermedades en general y de órganos internos de las mujeres; 4) Legislación: Normas, legislación, jurisprudencia, etc. que atañen a la mujer como persona; 5) Odontología y cirugía oral: Aspectos relacionados con el ejercicio de la odontología por mujeres; 6) Trabajo: Trabajo remunerado realizado por mujeres; 7) Derechos: Derechos de las mujeres a igualdad de estatus social, económico y de oportunidades educativas ofrecidas por la sociedad; 8) Servicios de salud: Servicios organizados para proporcionar atención de salud a la mujer; 9) Maltrato: Mujeres que son objeto de abuso físico y mental por un período de tiempo extenso, generalmente por parte del marido u otra figura masculina dominante.

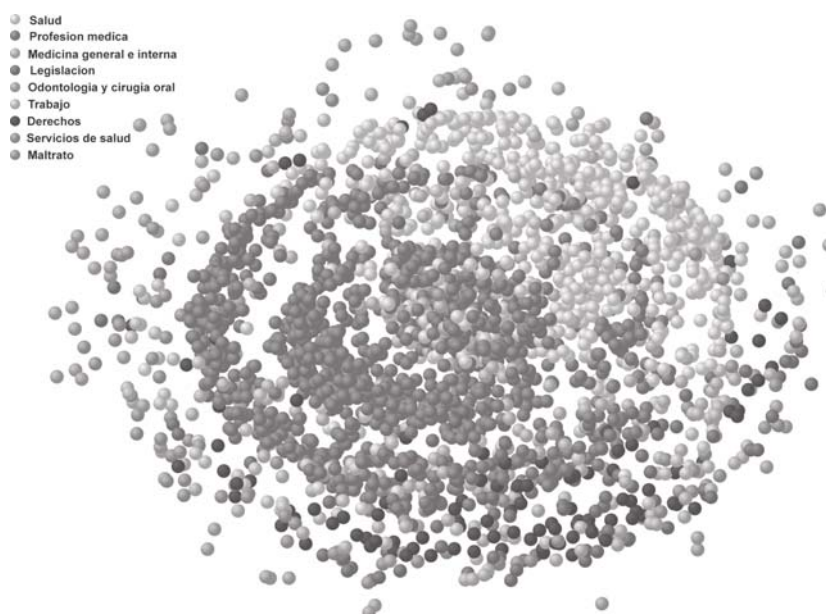
En la figura 2 se revela aún más el patrón centro-periferia al que antes hacíamos referencia. La concepción de una estructura formada por un centro y una periferia, constituye un paradigma clásico y recurrente en muchos campos de la ciencia (Bor-

Tabla I
Análisis factorial de la matriz de coocurrencia de revistas

<i>Factor</i>	<i>Nombre</i>	<i>Eigenvalue</i>	<i>% varianza</i>	<i>% acumulado de varianza</i>
1	Salud	1.909,1	72,1	72,1
2	Profesión médica	257,2	9,7	81,8
3	Medicina general e interna	135,7	5,1	86,9
4	Legislación	110	4,2	91,1
5	Odontología y cirugía oral	69,9	2,6	93,7
6	Trabajo	54,4	2,1	95,8
7	Derechos	33,5	1,3	97
8	Servicios de salud	26,7	1	98,1
9	Maltrato	24,9	0,9	99

gatti, 1999). Resulta revelador plasmar la idea clásica de la existencia de una estructura constituida por un núcleo activo en el que confluyen las principales áreas de investigación, formado por un entramado denso y compacto de revistas, frente a un conglomerado disperso en sus relaciones y poco conectado. En este caso, son dos fundamentalmente los factores o áreas de investigación que ocupan la zona central: *Salud y Profesión Médica*, poniendo así de manifiesto el alto grado de interés que prestan los investigadores a estas dos áreas. El resto se distribuye por la periferia, indicando un menor grado de implicación en la investigación relacionada con la mujer y su vinculación con las dos áreas principales.

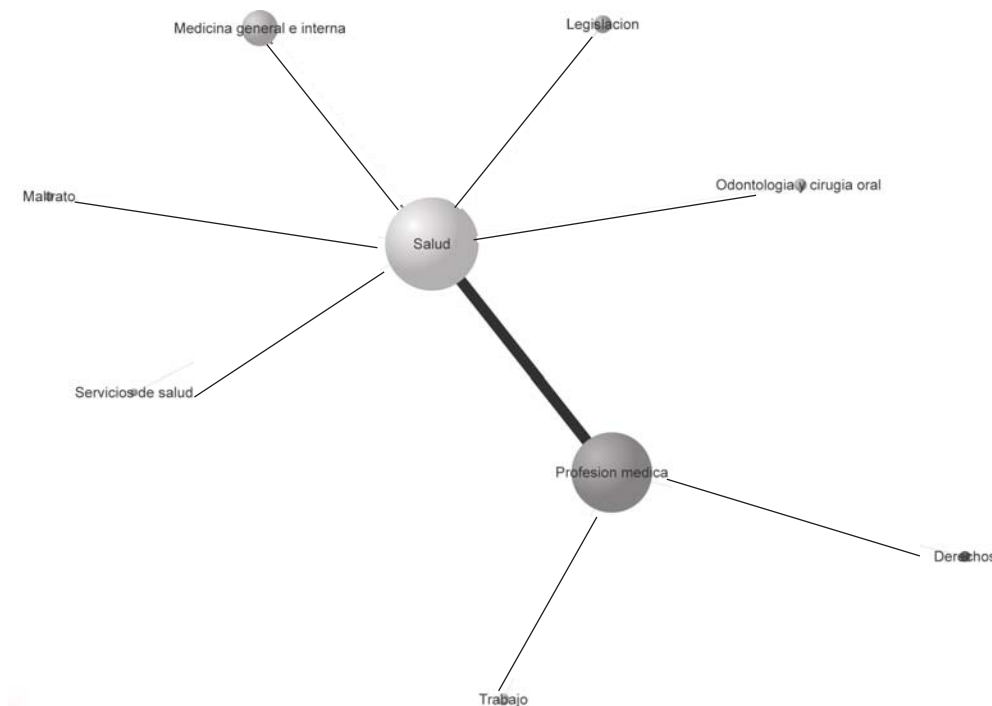
Figura 2
Grafo de coocurrencia de revistas en combinación con AF



Aglutinando las revistas de cada factor bajo su nombre correspondiente y estableciendo su grado de coocurrencia a partir de los descriptores que comparten cada uno de ellos, obtenemos un grafo muy clarificador que identifica perfectamente, a nivel macroestructural, las principales áreas de investigación relacionadas con la mujer (figura 3).

El nombre y color de cada área temática es el mismo que el de la imagen anterior, pero de esta forma se muestra mucho más clara, al aparecer las revistas agrupadas por áreas temáticas. Su tamaño es proporcional a la cantidad de documentos que en cada una de ellas se han publicado. Los enlaces indican las relaciones esen-

Figura 3
Principales áreas de investigación relacionadas con la mujer



ciales entre áreas, puesto que las más superfluas han sido eliminadas mediante PFNET, y su grosor denota el grado de relación entre las distintas áreas: a mayor grosor, mayor relación.

A simple vista observamos que las dos áreas de mayor tamaño, es decir, en las que más se investiga y que por tanto ocupan una posición más central, son: *Salud* y *Profesión Médica*, tal y como dedujimos de la figura 2.

Recurriendo al análisis de redes sociales (Wasserman, 1998), y en concreto, a las medidas de centralidad, podemos detectar además el grado de prominencia de cada una de las áreas temáticas. Estas medidas están destinadas a la detección e identificación de los actores centrales o más importantes de una red y se basan en la teoría de grafos. La idea de centralidad no hace referencia a la posición de un actor, sino a su grado de integración o cohesión en la red. Fundamentalmente podemos distinguir tres medidas de centralidad: grado, cercanía o *closeness*, y nivel de intermediación o *betweenness*. Un actor con un grado de centralidad elevado tendrá un amplio vecindario, ocupará posiciones centrales, se hará especialmente visible y se convertirá en un elemento importante para la interconexión de las distintas áreas temáticas de la red. Cuanto más elevada sea su centralidad por cercanía, mayor será su capacidad de interacción con el resto de áreas. Y por último, cuanto mayor sea su

poder de intermediación, más capacidad tendrá para actuar como punto de interconexión entre dos áreas temáticas. Para el caso que nos ocupa, aquellas áreas temáticas con una mayor centralidad de grado, de intermediación y de cercanía, serán consideradas las más prominentes.

En vista de los resultados de la tabla II, el área que más influencia tiene sobre el resto, como muestran sus enlaces, es *Salud*. Le sigue a cierta distancia *Profesión Médica*, y a continuación el resto de áreas con iguales valores y, por tanto, con un grado de participación similar en la investigación del dominio que nos ocupa.

Tabla II
Medidas de centralidad de grado

<i>Área de Investigación</i>	<i>Grado</i>	<i>Intermediación</i>	<i>Cercanía</i>
Salud	0,75	0,89	0,80
Profesión médica	0,375	0,46	0,61
Medicina general e interna	0,125	0	0,47
Legislación	0,125	0	0,47
Odontología y cirugía oral	0,125	0	0,47
Servicios de salud	0,125	0	0,47
Maltrato	0,125	0	0,47
Trabajo	0,125	0	0,40
Derechos	0,125	0	0,40

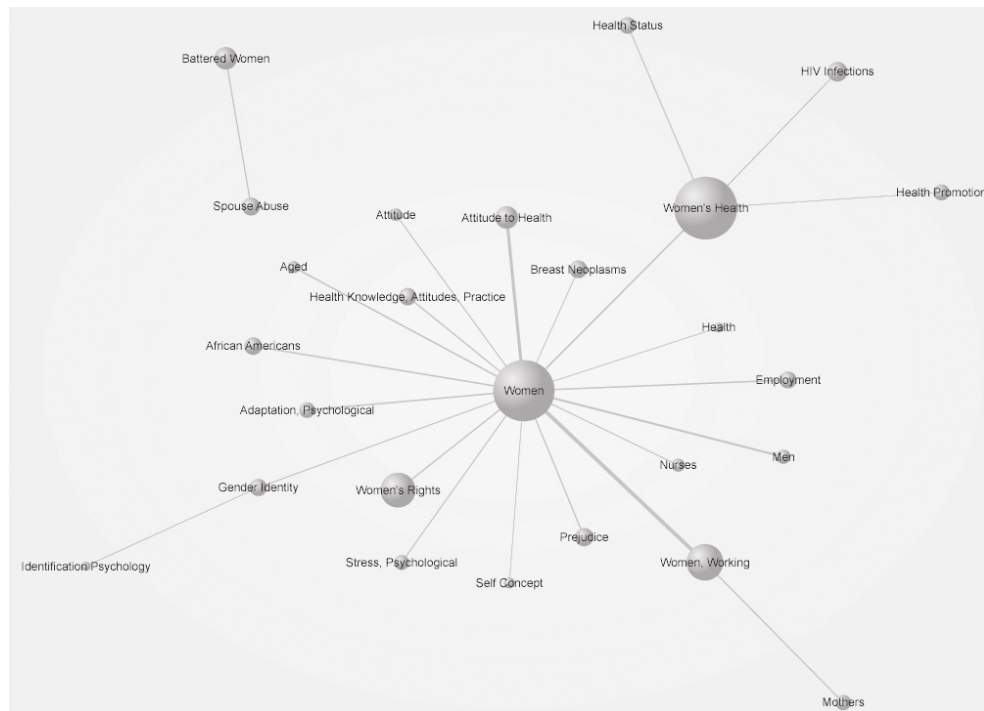
Para descender o ahondar en la estructura de cualquiera de estas áreas temáticas, sólo es necesario representar aquellos descriptores con un mayor grado de coocurrencia en cada área temática. Y es lo que hacemos en la figura 4 con el área *Salud*.

Al igual que en los casos anteriores, el tamaño del nodo o esfera que representa a cada descriptor, se ha hecho proporcional a la cantidad de documentos que contiene dicho descriptor. Su color es el mismo que el del área de investigación de la que procede. Los enlaces entre descriptores son los más significativos, habiendo eliminado los más superfluos con PFNET, y su grosor denota el grado de coocurrencia entre unos y otros.

Sin necesidad de recurrir a las medidas de centralidad, podemos deducir que el descriptor que más interviene en la investigación en esta área es *Women*. A su alrededor podemos observar el resto de descriptores sobre los que ejerce una mayor influencia. Estos indican los distintos temas de investigación. Incluso, podemos detectar a simple vista, agrupaciones o componentes aislados que revelan nuevas líneas de investigación dentro de esta área, como son, por ejemplo, *Women's Health* y *Spouse Abuse*.

Finalmente, y para realizar un estudio más detallado y pormenorizado de la investigación en cada área, mostramos, de forma independiente, las relaciones más significativas de cada uno de los descriptores más prominentes dentro de cada área temática. Esto lo hacemos con un mapa heliocéntrico (Moya-Anegón, 2005), al más puro estilo Atlas de la Ciencia (Grupo SCImago, 2005), como se puede ver en la figura 5, al mostrar los nodos más coocurrentes con *Breast Neoplasms*.

Figura 4
Veintiséis descriptores más coocurrentes de Salud

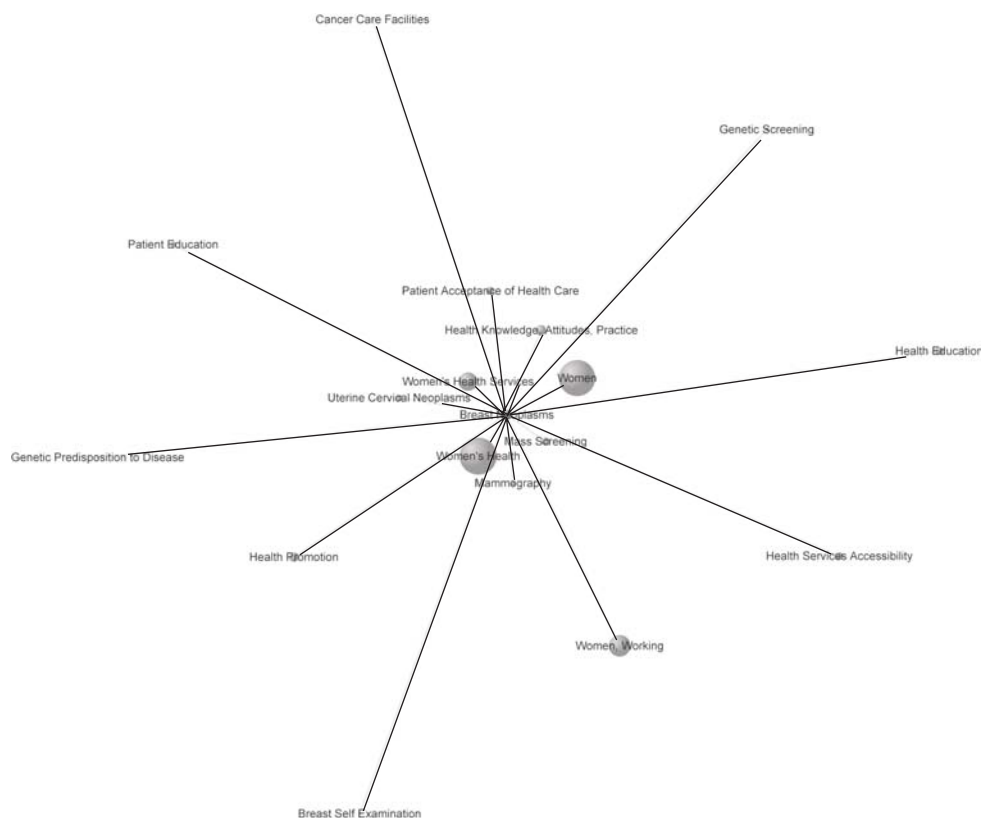


De nuevo, el tamaño de los nodos es proporcional a la cantidad de documentos publicados que contienen ese descriptor. Su color está determinado por el área a la que pertenece. Pero a diferencia de los mapas anteriores, el nivel de relación de unos descriptores con otros no depende del grosor de los enlaces, sino de su distancia, de ahí su nombre de mapa heliocéntrico. A mayor proximidad de un descriptor *satélite* con el descriptor central (*helios*), mayor será el número de documentos que contengan esos dos términos, y viceversa. Dichos documentos, están representados por el enlace que conecta a dos descriptores, y si se desea, su localización y acceso es tan fácil como lanzar una consulta del tipo: *descriptor heliocéntrico AND descriptor satélite*, contra el repositorio creado.

4 Conclusiones

Hemos desarrollado una metodología que, en tres niveles, es capaz de poner de manifiesto la estructura intelectual de la investigación relacionada con los aspectos médicos y clínicos de las mujeres en Medline. A través de las *palabras* de los indizadores de Medline, que son los encargados de definir las materias de investigación de los trabajos mediante el *lenguaje MeSH*, podemos ver las principales líneas de investi-

Figura 5
Mapa heliocéntrico de Breast Neoplasms



gación, especialidades y tendencias, ajustadas a la realidad de la producción científica almacenada en esta base de datos. El primer nivel consiste en una fase de acercamiento al tema que se pretende estudiar. Para ello se han recuperado todos los documentos en los que al menos aparece uno de los 9 términos que el DeCS utiliza para referirse al concepto «mujeres» (*women*). El segundo nivel pretende identificar los términos más relacionados con el tema objeto de estudio. Esto se consigue mediante la detección de los términos o «vecinos» más relacionados con los 9 DeCS de los que partíamos. El tercer y último nivel radica en la identificación temática del objeto de estudio, a partir de los documentos recuperados con los términos obtenidos en el segundo nivel.

En este caso se ha representado la investigación sobre un periodo de tiempo determinado. Pero este puede ser fraccionado con el fin de realizar estudios temporales y mostrar la evolución de la investigación.

En este trabajo sólo hemos representado *Cameos* de descriptores. Pero perfectamente pueden hacerse de revistas, autores y documentos, indicando así los más relevantes en cada área de investigación o materia.

Agradecimientos

Este trabajo ha sido posible gracias a la financiación pública obtenida mediante el *Contrato Art. 83 LOU. Ref. UAH 142/2005* del Observatorio de la Mujer del Ministerio de Sanidad y Consumo Español.

Bibliografía

- BATAGEL, J. V.; MRVAR, A. (2003). Analysis and visualization of large networks. En: Juenger, M.; Mutzel, P. (editores), *Graph Drawing Software*. Berlín: Springer.
- BATAGEL, J. V.; MRVAR, A. (2006). Pajek 1.11: Package for large network analysis.
- BIREME. DeCS: Descriptores en ciencias de la salud [página web]. Disponible en: <http://decs.bvs.br/E/decsweb.htm> [Consultado en: 12-12-2005].
- BORGATTI, S. P.; EVERETT, M. G. (1999). Models of Core/Periphery Structures. *Social Networks*, 21 (4), 375-395.
- BÖRNER, K.; CHEN, C.; BOYACK, K. W. (2003). Visualizing knowledge domains. *Annual Review of Information Science & Technology*, vol. 37, 179-255.
- BOYACK, K. W.; KLAVANS, R.; BÖRNER, K. (2005). Mapping the backbone of science. *Scientometrics*, 64 (3), 351-374.
- BRANDENBURG, F. J.; HIMSOLT, M.; ROHRER, C. (1996). An Experimental Comparison of Force-Directed and Randomized Graph Drawing Algorithms. *Lecture Notes in Computer Science*, 1027 (76), 87.
- CHEN, C. (1998). Generalised Similarity Analysis and Pathfinder Network Scaling. *Interacting with computers*, 10 (2), 107-128.
- CHEN, C. CiteSpace: visualizing patterns and trends in scientific literature [página web]. Disponible en: <http://cluster.cis.drexel.edu/~cchen/citespace/> [consultado en: 12-12-2005].
- DIRECCIÓN GENERAL DE SALUD PÚBLICA. La Declaración de Madrid. La equidad de Género en la salud mediante el mainstraming: la necesidad de avanzar [página web]. Disponible en: http://www.msc.es/Diseno/proteccionSalud/proteccion_todos_ciudadanos.htm [consultado en: 12-12-2005].
- FRUCHTERMAN, T.; REINGOLD, E. (1991). Graph Drawing by Force-Directed Placement. *Software Practice and Experience*, 21 (11), 1129-1164.
- GENDER AND HEALTH GROUP AT THE LIVERPOOL SCHOOL OF TROPICAL MEDICINE.(2000). Guidelines for the analysis of Gender and Health [página web]. Disponible en: <http://www.liv.ac.uk/lstm/hsr/GG-1.html> [consultado en: 8-02-2007].
- GRUPO SCIMAGO. (2005). Atlas of Science [página web]. Disponible en: www.atlasofscience.net [consultado en: 31-03-2005].
- KAMADA, T.; KAWAI, S. (1989). An algorithm for drawing general undirected graphs. *Information Processing Letters*, 31 (1), 7-15.
- LEYDESDORFF, L.; VAUGHAN, L. W.(2006). Co-occurrence matrices and their applications in information science: extending ACA to the Web environment. *Journal of the American Society for Information Science and Technology*, 52 (12), 1616-1628.
- LIN, X. ConceptLink [página web]. Disponible en: <http://faculty.cis.drexel.edu/~xlin/conceptlink.html> [consultado en: 12-12-2005].

- MOYA-ANEGÓN, F. D.; JIMÉNEZ CONTRERAS, E.; MONEDA CARROCHANO, M. D. L. (1998). Research fronts in library and information science in Spain (1985-1994). *Scientometrics*, 42 (2), 229-246.
- MOYA ANEGÓN, F. D.; VARGAS-QUESADA, B.; HERRERO-SOLANA, V.; CHINCHILLA-RODRÍGUEZ, Z.; CORERA-ÁLVAREZ, E.; MUNOZ-FERNÁNDEZ, F. J. (2004). A new technique for building maps of large scientific domains based on the cocitation of classes and categories. *Scientometrics*, 61 (1), 129-145.
- MOYA-ANEGÓN, F. D.; VARGAS-QUESADA, B.; CHINCHILLA-RODRÍGUEZ, Z.; HERRERO-SOLANA, V.; CORERA-ÁLVAREZ, E.; MUNOZ-FERNÁNDEZ, F. J. (2005). Domain analysis and information retrieval through the construction of heliocentric maps based on ISI-JCR category cocitation. *Information Processing & Management*, vol. 41 (6), 1520-1533.
- NIH Revitalization Act of 1993, Pub L No.103-43.
- NATIONAL LIBRARY OF MEDICINE. (2005a). Medline [página web]. Disponible en: <http://www.nlm.nih.gov/databases/> [consultado en: 12-12-2005].
- NATIONAL LIBRARY OF MEDICINE. (2005b). PubMed [página web]. Disponible en: <http://www.pubmed.gov> [consultado en: 12-12-2005].
- NATIONAL LIBRARY OF MEDICINE. (2006). Medical Subject Headings [página web]. Disponible en: <http://www.nlm.nih.gov/mesh/MBrowser.html> [consultado en: 12-12-2005].
- OVID TECHNOLOGIES. (2005). Ovid [Página Web]. Disponible en: <http://www.ovid.com> [consultado en: 10-02-2006]
- SAMOYLENKO, I.; CHAO, T.-C.; LIU, W.-C.; CHEN, C.-M. (2006). Visualizing the scientific world and its evolution. *Journal of the American Society for Information Science & Technology*, 57 (11), 1461-1469.
- SEN, G.; GEORGE, A.; ÖSTLIN, P. (2005). *Incorporación de la perspectiva de género en la investigación y en las políticas en salud: una perspectiva de género*. Harvard: Harvard Center for Population and Development Studies / Organización Panamericana de la Salud.
- W3C (2004). Scalable Vector Graphics (SVG) 1.1 Specification [página web]. Disponible en: <http://www.w3.org/TR/SVG11/> [consultado en: 10-01-2006].
- WASSERMAN, S.; FAUST, K. (1998). *Social network analysis: methods and applications*. Cambridge: University Press.
- WHITE, H. D. (2003). Pathfinder networks and author cocitation analysis: a remapping of paradigmatic information scientist. *Journal of the American Society for Information Science and Technology (JASIST)*, 54 (5), 423-434.
- WIZEMANN T. M.; PARDUE M. L. (2001). *Exploring the Biological Contributions to Human Health: Does Sex Matter?* Washington, DC: Board on Health Sciences Policy, Institute of Medicine.
- WHO. WORLD HEALTH ORGANIZATION (2002). Gender Policy for WHO [página web]. Disponible en: <http://www.who.int/gender/documents/sppolicy.pdf> [consultado en: 08-02-2007].
- Women's Health Bureau.(2002). Exploring concepts of gender and health [página web]. Disponible en: <http://www.hcsc.gc.ca/english/women/exploringconcepts.htm> [consultado en: 08-02-2007].